JOINT SUCCESSIVE CORRELATION ESTIMATION AND SIDE INFORMATION REFINEMENT IN DISTRIBUTED VIDEO CODING

Nikos Deligiannis[a,c], Frederik Verbist[a,c], Jürgen Slowack[b,c], Rik Van de Walle[b,c], Peter Schelkens[a,c], and Adrian Munteanu[a,c]

[a]Department of Electronics and Informatics, Vrije Universiteit Brussel, Pleinlaan 2, B-1050 Brussels, Belgium.
[b]ELIS Department, Multimedia Lab, Ghent University, Gaston Crommenlaan 8, B-9050 Ghent, Belgium.
[c]Interdisciplinary Institute for Broadband Technology (IBBT), Gaston Crommenlaan 8, B-9050 Ghent, Belgium.

ABSTRACT

This paper presents a novel hash-based distributed video coding (DVC) scheme that combines an accurate online correlation channel estimation (CCE) algorithm with an efficient side information refinement strategy, delivering state-of-the-art compression performance. The proposed DVC scheme applies layered bit-plane Wyner-Ziv coding and successively refines the CCE bit-plane-per-bit-plane during decoding. In addition, the side information is successively refined upon decoding of distinct refinement levels, grouping specific frequency bands of the discrete cosine transform. The proposed system not only outperforms the benchmark in DVC but several state-of-the-art side information refinement techniques and CCE methods as well.

Index Terms—Distributed video coding, correlation channel estimation, successive refinement

1. INTRODUCTION

Two main challenges in distributed video coding (DVC), also known as Wyner-Ziv (WZ) video coding, are (i) side information (SI) generation at the decoder by means of motion-compensated prediction and (ii) accurate capturing of the conditional dependency between the original source and its SI.

Common DVC schemes [1], [2] employ motion-compensated interpolation (MCI) to generate SI. Alas, the compression performance of MCI-based DVC architectures [2] deteriorates when the amount of motion in the sequence or the GOP size increases [2]. To improve the quality of the created SI, alternative DVC schemes employ hash-based motion estimation [1], [3] where proper hash information is transmitted to the decoder to support motion prediction. Additionally, joint decoding and SI refinement (SIR) has been shown to improve the compression performance.

In transform-domain WZ (TDWZ) systems, a post-processing procedure, including motion vector refinement and optimal reference frame selection, to update the SI prior to the final coefficient reconstruction was proposed in [4]. Alternatively, repeated SI generation after the reconstruction of every discrete cosine transform (DCT) coefficient band was put forward in [5], where only the already available SI frame was used as a reference to further refine the SI. Similar to [5], reference information for SIR was extracted from the previous and the next reference frames in [6]. Performing SIR upon DC band decoding only was used in [7] without the need for an overcomplete DCT.

Regarding the second challenge in DVC, accurate modeling and correlation channel estimation (CCE) is essential. The majority of practical DVC solutions assume an additive noise channel, where the correlation noise (CN) is an independent memoryless zero-mean Laplacian distributed random variable [1]. Various stationarity levels of the CN have been considered. In TDWZ, the CN can be estimated for every frequency band per frame, as was performed online in [8]. Adjusting the CN model per transform coefficient, comprising the finest estimation level, was carried out online in [8] as well. Alternatively, in [9] the parameters of the CN in the transform-domain were derived by exploiting a pixel-domain CCE and the spatial correlation of the noise signal. An analogous concept was employed by the TRACE algorithm in [10], where progressive refinement of the noise variance upon decoding of each DCT band was carried out. A particle filtering approach to CCE that exploits the inter-bit-plane correlation in a TDWZ architecture, was proposed in [11]. The notion of SI-dependent (SID) CN was introduced in [12]. In [3], the SID noise model was utilized in a hash-based DVC (HDVC) architecture along with an efficient online band level bit-plane-by-bit-plane successively refined CCE. It was theoretically and experimentally shown that an SID correlation channel improves the compression performance with respect to its SI-independent (SII) counterpart [3].

Advancing over [3], this work presents a novel HDVC system that marries both the concepts of online successively refined SID CCE and SIR, yielding state-of-the-art TDWZ compression performance. The proposed HDVC creates SI by means of hash-based overlapped block motion estimation and compensation (OBMEC) [3] which is successively refined over so-called refinement levels (RLs), consisting of a number of grouped DCT frequency bands. In contrast to [5], [6], the proposed technique derives high-quality SI with a limited number of RLs. Nested inside the SIR loop, an SID CCE is progressively refined per additional decoded bit-plane. The proposed SID CCE approach is not confined to a specific SI generation method, unlike the CCE

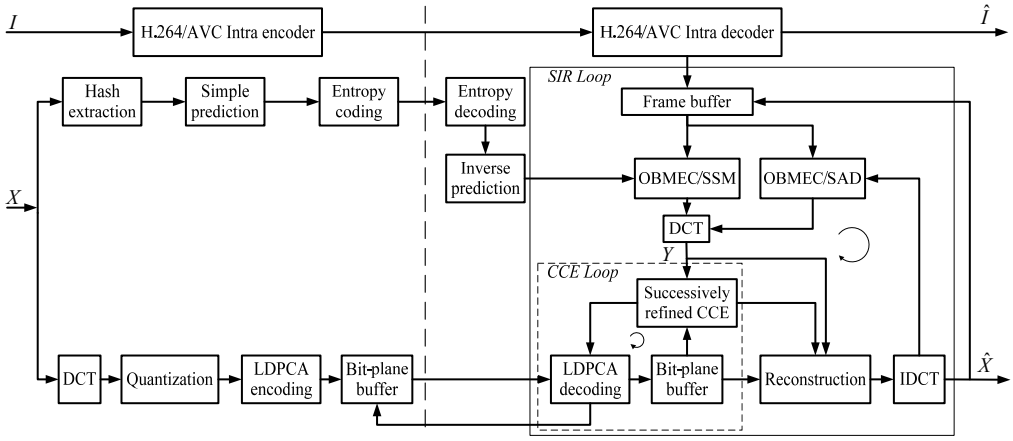

Fig. 1. The proposed HDVC architecture supporting SIR.

technique from [8], which is limited to MCI-based DVC. Moreover, the presented SID CCE does not impose a fixed decoding order of the DCT frequency bands, unlike the TRACE [10] algorithm which impedes the decoding order of the proposed RLs. Experimental results corroborate the accuracy of the proposed CCE between the source and the successively refined SI, which outperforms state-of-the-art estimators [8], [10]. Moreover, with respect to alternative state-of-the-art DVC schemes, the proposed HDVC delivers superior compression performance.

The rest of the paper is organized as follows. Section 2 offers an overview of the proposed DVC architecture. Section 3 covers the proposed CCE algorithm, while Section 4 focuses on the SIR strategy. Section 5 provides experimental results and Section 5 concludes the paper.

2. SYSTEM ARCHITECTURE

The HDVC architecture, presented in this paper, is depicted in Fig. 1. At the encoder, the input sequence is split into key and WZ frames, which are organized in groups of pictures (GOPs) and coded similar to [3]. Namely, the key frames are coded using H.264/AVC Intra while the WZ frames undergo a 4×4 integer DCT followed by quantization and Slepian-Wolf (SW) coding, based on rate-adaptive low-density parity-check accumulate (LDPCA) codes, employing a feedback channel. The proposed system forms an additional hash for every WZ frame, according to the procedure in [3], i.e., the hash consists of the most significant bit-plane of the downscaled WZ frames, which is spatially decorrelated and entropy coded.

At the decoder, the key frames and hash information are decoded and stored for SI generation. Each WZ frame is decoded in distinct stages, referred to as RLs. Every such RL comprises SI generation, SID CCE and WZ decoding of a specific set of DCT coefficient bands of the frame. Thus, starting from the hash information, the decoder gains more knowledge of the original WZ frame as more RLs are decoded. In this way, the quality of the SI generated at each RL is consecutively improved (see *SIR loop* in Fig. 1).

Within each RL, online SID CCE is executed for every frequency band belonging to that specific RL. Per band, the decoder produces soft-input from the CCE and each SI coefficient to decode the WZ bit-planes. Per decoded bit-plane of a band, the algorithm is executed again, thus enabling bit-plane-by-bit-plane progressively refined SID CCE (see *CCE loop* in Fig. 1). Hence, the presented HDVC performs consists of two nested refinement loops, refining both the SI and the CCE.

3. PROGRESSIVE SID CCE

The presented progressive SID CCE algorithm is compatible with a WZ architecture following a layered bit-plane approach. Let X, Y be the random variables representing the coefficients of a given frequency band in a specific RL in the original WZ and SI frame, respectively. We consider an additive noise channel, $X = Y + N$, where N denotes the CN random variable. Adhering to an SID [3], [12] correlation channel concept, the dependency of the channel output X on the input Y is expressed by a conditional probability density function (pdf) $f_{X|Y}\left(x|y\right)$, which is a Laplacian distribution centered on the realization y of the SI and of which the standard-deviation $\sigma\left(y\right)$ varies depending on the particular realization y. For the sake of practicality [3], a finite number of standard-deviations $\sigma\left(y_k\right)$ are considered. In this way, every single $\sigma\left(y_k\right)$ is assigned to all SI values indexed by $y_k \in A_y = \{y_0, y_1, \ldots, y_K\}$ via quantization. Suppose that the coefficients in the band have been quantized using an L-bit quantizer. The progressive refinement algorithm is initiated after SW decoding of the first bit-plane of the band. Any online CCE method, e.g., [8], [9] can be employed to this end. In our implementation, the initial $\sigma\left(y_k\right)$ estimates are taken from the co-located frequency bands in the previously decoded WZ frame and the corresponding SI.

The proposed online algorithm, which is sketched in Table 1, uses the previously decoded bit-planes to refine the CCE for decoding the next bit-plane, hence enabling bit-plane-by-bit-plane progressive refinement. Specifically, let the binary tuples $\mathbf{b}_m$, $1 \le m \le L$ represent the already SW decoded bit-planes of the coefficients in the band. The

TABLE 1. THE PROGRESSIVE SID CCE ALGORITHM.

Initialize $\mathbf{q}_1$

From $m = 1,2,\ldots,L$ do:

1) Determine the joint pmf $p_{Q_m,Y}\left(q_{m,\vartheta}, y_k\right)$ by normalizing the histogram of joint observations $\left(q_{m,\vartheta}, y_k\right)$, $\forall \vartheta \in \left[0, 2^m - 1\right]$ and $\forall y_k \in A_y$
2) Derive the pmf $p_Y\left(y_k\right) = \sum_{\vartheta=0}^{2^m-1} p_{Q_m,Y}\left(q_{m,\vartheta}, y_k\right)$
3) Calculate the empirical conditional pmf according to: $p_{Q_m|Y}\left(q_{m,\vartheta}|y_k\right) = p_{Q_m,Y}\left(q_{m,\vartheta}, y_k\right) / p_Y\left(y_k\right)$
4) Find $\sigma\left(y_k\right)$ such that:

$$p_{Q_m|Y}\left(q_{m,\vartheta}|y_k\right) - \int_{q_L}^{q_H} f_{X|Y}\left(x|y\right)dx = 0\ ,\ \forall y_k \in A_y\ , \quad (1)$$

where q_L and q_H are the upper and lower bound of the quantization interval corresponding to bin $q_{m,\vartheta}$. The derivation of the uniqueness of the root of (1) is found in [3]

5) If ($m < L$)
   use $f_{X|Y}\left(x|y\right)$ to generate the soft input to LDPCA decode the next bit-plane $\mathbf{b}_{m+1}$
   Else
   perform reconstruction

algorithm combines these bit-planes to derive intermediate quantization indices $\mathbf{q}_m$, that constitute a coarse version of the band's coefficients. Namely, the tuple $\mathbf{q}_m$ demarcates 2^m quantization bins $q_{m,\vartheta}$, enumerated by $\vartheta \in \left[0, 2^m - 1\right]$. Then, based on the available partial indices and the discretized SI coefficients of the band, the conditional probability mass function (pmf) $p_{Q_m|Y}\left(q_{m,\vartheta}|y_k\right)$ is empirically generated for any given $y_k \in A_y$, as instructed in Table 1. From the obtained pmf per y_k, the algorithm derives the matching conditional pdf. Intuitively, per y_k, the algorithm searches for the conditional pdf $f_{X|Y}\left(x|y\right)$–defined by $\sigma\left(y_k\right)$–that generates a conditional pmf matching the empirically obtained $p_{Q_m|Y}\left(q_{m,\vartheta}|y_k\right)$. This is achieved by solving Eq. (1) given in Table 1. The resulting pdf $f_{X|Y}\left(x|y\right)$ is then used to generate the soft-input to decode the next bit-plane.

As the number of already decoded bit-planes m increases a finer representation of the source is available to the CCE algorithm and consequently $p_{Q_m|Y}\left(q_{m,\vartheta}|y_k\right)$ will more closely match the true correlation channel characteristics. This property of following a layered WZ coding approach is exploited by consecutively updating the model parameters after decoding every new bit-plane $\mathbf{b}_{m+1}$. When all L bit-planes have been decoded, the algorithm is executed again and the source samples are optimally reconstructed.

We remark that the accuracy of the proposed SID CCE algorithm is tuned by the assumed noise stationarity level and the number of quantization levels (QLs) K of the SI coefficients. Concerning stationarity, the presented CCE can be carried out on any given collection of spatially neighboring coefficients in the band. Although regulating the CCE on small areas adjusts the noise component to the regional variation of the channel statistics, superior performance is not guaranteed since the number of samples for statistical inference diminishes. Analogously, although increasing the number of QLs K captures the SID model better, online CCE might suffer from lack of statistical support. Experimental evidence has shown that, for the considered frame resolutions, applying CCE at the band level and quantizing the SI using the same number of QLs per band as for the WZ frame, yields the best performance.

Refinement Level	DCT coefficient band(s)
RL_0	0
RL_1	1, 2
RL_2	3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15

0	1	5	6
2	4	7	12
3	8	11	13
9	10	14	15

Fig. 2. Definition of the frequency bands per RL.

4. SIDE INFORMATION REFINEMENT

To enable SIR, a WZ frame is decoded in several RLs. The higher the number of RLs the more computational complexity due to repeated SI generation is imposed on the decoder. On the other hand, since the SI quality is progressively improved, the overall WZ rate and thereby the LDPCA soft-decoding complexity are reduced. In contrast to [5], [6] where block-based motion estimation was used during the refinement stages, the proposed system performs OBME. With respect to block-based techniques, OBME enables multi-hypothesis pixel-based motion-compensation producing high-quality SI at the cost of additional computations. Hence, to balance the aforementioned effects, the number of RLs in our scheme is set to three. In particular, every RL is built around specific frequency bands of the 4×4 DCT, which are assigned to the different RLs as shown in Fig. 2. The considered grouping is inspired by the distribution of the frequencies in the 4×4 DCT block.

In the first level, i.e., RL_0, the decoder uses the hash and already decoded key and/or WZ frames to create the initial SI Y_0 by means of OBMEC with sub-sampled matching (OBMEC/SSM) [3]. Because of the binary nature of the hash, motion search is executed using the Hamming distance as an error metric. This initial SI is employed to decode the DC band of the WZ frame. The presented online progressive SID CCE algorithm is performed as in Section 3 based on which SW decoding and reconstruction of the DC coefficient band are carried out. Finally, the reconstructed DC coefficients and the coefficients of the SI Y_0 at the positions of the remaining RLs are congregated and the inverse DCT is performed, yielding the partially decoded WZ frame $\hat{X}_0$.

The decoder then enters RL_1, the next SIR stage, where the decoder now has the partially decoded WZ frame $\hat{X}_0$ and hence more knowledge of the original WZ frame at its disposal. This is exploited by performing OBMEC using $\hat{X}_0$ and the reference frames to generate the refined SI Y_1. Remark that conversely to the OBMEC/SSM method used in RL_0, where motion estimation was carried out based on the hash, block matching is now performed at the original

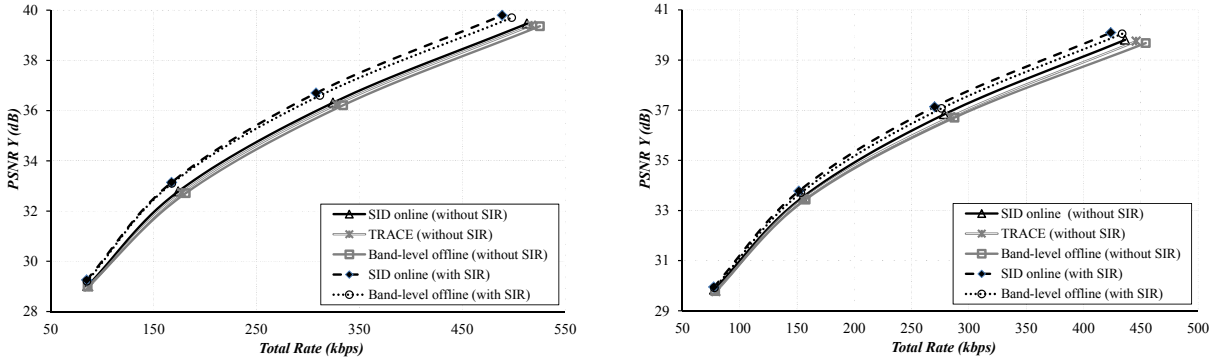

Fig. 3. Compression results of the proposed HDVC using different CCE algorithms for the entire (a) Foreman and (b) Carphone sequences at QCIF resolution, a frame rate of 15Hz and in a GOP of 4.

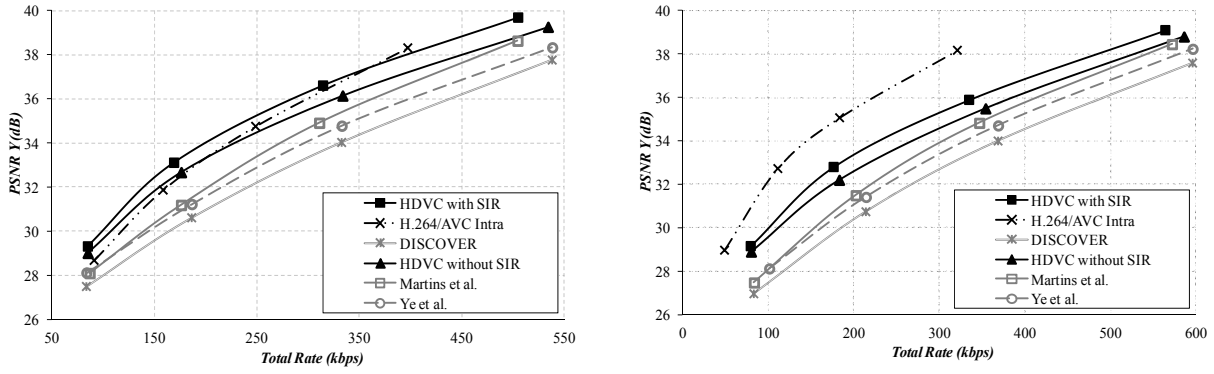

Fig. 4. Comparative compression results of the proposed HDVC with SIR for the entire (a) Foreman and (b) Soccer sequences at QCIF resolution, a frame rate of 15Hz and in a GOP of 8.

frame resolution using the traditional sum of absolute differences (SAD) error metric.

Given the updated SI Y_1, LPDCA decoding and reconstruction is carried out to obtain the reconstructed coefficients of every band in RL_1. The required soft information is extracted from the proposed bit-plane-by-bit-plane progressive SID CCE. It is important to note that although the previous stage RL_0 has already been decoded, the improved SI Y_1 is still useful to improve the quality of the reconstruction of the frequency band in RL_0. To this end, SID CCE and reconstruction are executed again for the DC band comprising RL_0 using the updated SI Y_1 and the already decoded bit-planes. Before entering the final RL, all updated frequency bands are gathered with the SI coefficients (belonging to still not decoded frequencies) and by inverse DCT the partially decoded WZ frame $\hat{X}_1$ is obtained.

The same SI generation process and WZ decoding is executed again in the last round of refinement, i.e., RL_2, yielding the final reconstructed WZ frame $\hat{X} = \hat{X}_2$.

4. EXPERIMENTAL RESULTS

A first set of experiments illustrates the benefit of combining the proposed SID CCE and SIR approach in the presented HDVC architecture. To show the effect of our SIR technique, results of the proposed HDVC architecture without SIR are also presented. Both versions of our HDVC, i.e., with and without SIR, were equipped with the proposed online progressive SID CCE as well as the offline band-level SII CCE algorithm in [8]. The HDVC version without SIR additionally features the online SII coefficient-level channel estimator TRACE [10], which could not be integrated in the version supporting SIR since the fixed order in which the DCT frequency bands need to be decoded interferes with the pursued SIR strategy.

The compression performance of the different configurations of the proposed HDVC system is shown in Fig. 3(a) and (b) for the Foreman and Carphone sequence, respectively. The results reveal that adopting the SID CCE in our HDVC without SIR increases the compression performance compared to when the TRACE [10] algorithm or the offline SII band level method of [8] are used, with respective Bjøntegaard [13] rate savings of 2.20% and 4.19% for Foreman and 2.50% and 4.07% for Carphone. Moreover, adding our proposed SIR technique significantly raises the average PSNR compared to [3], irrespective of which CCE algorithm is used. Although the increase in SI quality reduces the required rate for successful LPDCA decoding, and thereby the potential impact of CCE, the proposed progressive SI CCE still brings Bjøntegaard rate savings of 1.69% and 2.51% over the offline SII CCE in [8] for Foreman and Carphone, respectively.

TABLE 2. BJØNTEGAARD [13] RATE SAVINGS (%) AND PSNR REDUCTION (DB) OF THE PROPOSED HDVC WITH SIR RELATIVE TO ALTERNATIVE DVC SCHEMES.

		GOP2		GOP4		GOP8	
		ΔR	ΔPSNR	ΔR	ΔPSNR	ΔR	ΔPSNR
a)	DISCOVER	-9.01	0.584	-26.20	1.712	-38.78	2.747
	Ye et al.	-6.98	0.470	-19.03	1.209	-31.04	2.102
	Martins et al.	-1.65	0.116	-14.57	0.965	-25.58	1.753
b)	DISCOVER	-18.09	1.028	-31.15	1.976	-39.38	2.685
	Ye et al.	-11.87	0.669	-24.24	1.481	-31.12	2.024
	Martins et al.	-7.55	0.439	-20.97	1.284	-25.63	1.671

In a second set of experiments, the compression performance of the proposed HDVC with and without SIR is compared against the DISCOVER and H.264/AVC Intra codecs, serving as benchmark, as well as against the alternative DVC solutions in [4] and [5] that also successively refine the SI. The experimental results for the Foreman and Soccer sequence in a GOP of 8 are shown in Fig. 4 (a) and (b), respectively. The compression gains of our HDVC with SID CCE and the proposed SIR in terms of Bjøntegaard rate savings and Bjøntegaard PSNR increase with respect to DISCOVER and the SIR DVC systems in [4] and [5], are summarized in Table 2 for (a) Foreman and (b) Soccer in a GOP of 2, 4 and 8. These results confirm the superior compression performance of the proposed DVC architecture, with Bjøntegaard rate savings of up to 39.38% with respect to the benchmark DISCOVER codec and up to 31.12% and 25.63% compared to the scheme of [4] and [5], respectively, for Soccer in a GOP of 8.

5. CONCLUSIONS

This paper introduced a novel DVC architecture that incorporates online bit-plane-by-bit-plane successively refining SID CCE in conjunction with an efficient SIR strategy, based on RLs comprised of grouped DCT frequency bands. Experimental results show increased compression performance due to the proposed SID CCE with respect to SII based offline CCE, both when SIR is activated or disabled. Moreover, the evaluation of the proposed DVC architecture with SIR confirms state-of-the-art compression performance that significantly advances over the standard reference DISCOVER codec as well as several alternative DVC schemes featuring SIR.

6. ACKNOWLEDGEMENTS

This work was supported by the FWO Flanders projects G.0391.07, G.0146.10 and the postdoctoral fellowship of Peter Schelkens.

7. REFERENCES

[1] B. Girod, A. Aaron, S. Rane, and D. Rebollo-Monedero, "Distributed video coding," *Proceedings of the IEEE,* vol. 93, pp. 71-83, January 2005.

[2] X. Artigas, J. Ascenso, M. Dalai, S. Klomp, D. Kubasov, and M. Quaret, "The DISCOVER codec: Architecture, techniques and evaluation," in *Picture Coding Symposium, PCS*, Lisboa, Portugal, November 2007 [Online]. Available: www.discoverdvc.org.

[3] N. Deligiannis, J. Barbarien, M. Jacobs, A. Munteanu, A. Skodras, and P. Schelkens, "Side-information dependent correlation channel estimation in hash-based distributed video coding," *IEEE Transactions on Image Processing, accepted,* December 2011.

[4] S. Ye, M. Ouaret, F. Dufaux, and T. Ebrahimi., "Improved side information generation for distributed video coding by exploiting spatial and temporal correlations," *EURASIP Journal on Image and Video Processing,* pp. 1-15, January 2009.

[5] R. Martins, C. Brites, J. Ascenso, and F. Pereira, "Refining side information for improved transform domain Wyner-Ziv video coding," *IEEE Transactions on Circuits and Systems for Video Technology,* vol. 19, pp. 1327-1341, September 2009.

[6] A. Abou-Elailah, F. Dufaux, J. Farah, M. Cagnazzo, and B. Pesquet-Popescu, "Successive refinement of motion compensated interpolation for transform-domain distributed video coding," in *European Signal Processing Conference, EUSIPCO*, Barcelona, Spain, August-September 2011.

[7] N. Deligiannis, M. Jacobs, J. Barbarien, F. Verbist, J. Škorupa, R. Van de Walle, A. Skodras, P. Schelkens, and A. Munteanu, "Joint DC coefficient band decoding and motion estimation in Wyner-Ziv video coding," in *IEEE International Conference on Digital Signal Processing, DSP*, Corfu, Greece, July 2011.

[8] C. Brites and F. Pereira, "Correlation noise modeling for efficient pixel and transform domain Wyner-Ziv video coding," *IEEE Transactions on Circuits and Systems for Video Technology,* vol. 18, pp. 1177-1190, September 2008.

[9] J. Škorupa, J. Slowack, S. Mys, N. Deligiannis, J. De Cock, P. Lambert, A. Munteanu, and R. Van de Walle, "Exploiting quantization and spatial correlation in virtual-noise modeling for distributed video coding," *Signal Processing: Image Communication,* vol. 25, pp. 674 - 686, October 2010.

[10] X. Fan, O. C. Au, and N. M. Cheung, "Transform-domain adaptive correlation estimation (TRACE) for Wyner-Ziv video coding," *IEEE Transactions on Circuits and Systems for Video Technology,* vol. 20, pp. 1423-1436, November 2010.

[11] S. Wang, L. Cui, L. Stankovic, V. Stankovic, and S. Cheng, "Adaptive correlation estimation with particle filtering for distributed video coding," *IEEE Transactions on Circuits and Systems for Video Technology, to appear,* 2012.

[12] N. Deligiannis, A. Munteanu, T. Clerckx, J. Cornelis, and P. Schelkens, "On the side-information dependency of the temporal correlation in Wyner-Ziv video coding," in *IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP*, pp. 709–712, Taipei, Taiwan, April 2009.

[13] G. Bjontegaard, "Calculation of average PSNR differences between RD-curves," ITU-T Video Coding Experts Group (VCEG), Austin, TX, Document VCEG-M33, April 2001.